AD-A094 728 MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB F/G 12/1
ON NONPARAMETRIC PROBABILITY DENSITY ESTIMATION USING ORTHOGONA--ETC(U)
NOV 80 L K JONES F19628-80-C-0002
UNCLASSIFIED TN-1980-54 ESD-TR-80-230 NL

1 OF 1
AD
AD 94 728

END
DATE
FILMED
3-81
DTIC

AD A095728

Dec 14 73

(12)

LEVEL II

Technical Note	1980-54

L. K. Jones

DTIC
ELECTE
FEB 0 9 1981
S D
E

On Nonparametric Probability Density Estimation Using Orthogonal Series

26 November 1980

Prepared for the Department of the Air Force
under Electronic Systems Division Contract F19628-80-C-0002 by

Lincoln Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LEXINGTON, MASSACHUSETTS

Approved for public release; distribution unlimited.

DDC FILE COPY

81 2 09 006

The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, with the support of the Department of the Air Force under Contract F19628-80-C-0002.

This report may be reproduced to satisfy needs of U.S. Government agencies.

The views and conclusions contained in this document are those of the contractor and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the United States Government.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Raymond L. Loiselle

Raymond L. Loiselle, Lt. Col., USAF
Chief, ESD Lincoln Laboratory Project Office

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

ON NONPARAMETRIC PROBABILITY DENSITY ESTIMATION USING ORTHOGONAL SERIES

L. K. JONES

Group 92

Accession For
NTIS GRA&I X
DTIC TAB
Unannounced
A

TECHNICAL NOTE 1980-54

26 NOVEMBER 1980

Approved for public release; distribution unlimited.

LEXINGTON MASSACHUSETTS

ABSTRACT

A method of density estimation is proposed, which is a rational modification of orthogonal expansions, combined with a stopping rule determined by a nearest neighbor statistic. This method yields consistent estimates and applies (in principle) to density estimation in any number of dimensions.

ON NONPARAMETRIC PROBABILITY DENSITY ESTIMATION USING ORTHOGONAL SERIES

I. INTRODUCTION

Among the numerous non-parametric methods of estimating a probability density function, the approximation of this density by a finite fourier series has several computational advantages. Probably the most important of these is the fact that the evaluation of this density at a new data point requires only the storage of certain fourier coefficients. One of the main disadvantages of such an approximation of a density is the difficulty of determining the number of terms in the expansion.

In this note, we propose an approximation which is a rational function of a finite fourier series. The number of terms in this series depends, in a very natural way, on the nearest neighbor error rate for the sample data when compared to a sample drawn from a reference distribution. In III we show that the method is consistent and in IV we remark on the relevance of this method in hypothesis testing.

II. SECOND ORDER SOLUTION TO THE BINARY DECISION PROBLEM

Let p_1, p_2 be two Lebesgue measurable, bounded ($\leq K$) probability density functions on the unit cube, I, in R^n. We assume further that $p_1 \neq p_2$ on some set of positive measure in I and that for some $\delta > 0$, $p_1 \geq \delta$ on I. Let $\mathcal{L} = \{f \in L_2(I): E_1 f = \int f\, p_1 dx = 0,\ E_2 f = \int f\, p_2 dx = 1\}$. According to [1], a second order solution for an optimal discriminant $\overline{f} \in \mathcal{L}$, for the binary hypothesis test H_1: X has density p_1 vs. H_2: X has density p_2, is a critical point for some real α of the functional

$$J_\alpha(f) = \alpha \mathrm{VAR}_1\, f + (1-\alpha)\mathrm{VAR}_2\, f \tag{1}$$

In fact if we restrict ourselves to the case $0 < \alpha < 1$ and solve (1) for the unique (to within a null function) critical point (and global minimum of $J_\alpha(f)$ for $f \in \mathcal{L}$), we obtain by elementary variational calculus

$$\overline{f} = \frac{\left[(1-\alpha) - \lambda\right] p_2/p_1 + \lambda}{\alpha + (1-\alpha)\; p_2/p_1} \tag{2}$$

with

$$0 > \lambda = \frac{(1-\alpha)\displaystyle\int \frac{p_2 p_1}{\alpha p_1 + (1-\alpha)p_2}\, dx}{\displaystyle\int \frac{(p_2 - p_1)p_1}{\alpha p_1 + (1-\alpha)p_2}\, dx} = (1-\alpha) - J_\alpha(\overline{f})$$

It follows that $\overline{f}$ is rational and increasing in (p_2/p_1) and hence optimal (by an adjustment of threshold) for minimum total error (or Neyman-Pearson at level β) hypothesis testing.

III. DENSITY ESTIMATION

For simplicity we consider only the case $\alpha=\frac{1}{2}$. Similar results may be obtained for other α. Solving (2) for p_2/p_1 we obtain

$$\frac{p_2}{p_1} = \frac{\frac{1}{2}\bar{f}-\lambda}{(\frac{1}{2}-\lambda)-\frac{1}{2}\bar{f}} = \frac{\bar{f}-1+2J_{\frac{1}{2}}(\bar{f})}{2J_{\frac{1}{2}}(\bar{f})-\bar{f}} \tag{3}$$

We now write

$$J_{\frac{1}{2}}(\bar{f}) = \frac{1}{2} + \frac{\varepsilon_{nn}}{4(\frac{1}{2}-\varepsilon_{nn})} \tag{4}$$

where $\varepsilon_{nn} = \int \frac{p_2 p_1}{p_1+p_2}dx$ is known as the limiting nearest neighbor error rate, i.e.,if we generate n independent class 1 samples from a distribution with density p_1 and similarly n class 2 samples from a distribution with density p_2 and then classify new samples (drawn from class 1 or class 2 with equal probability) as the class of the (a) nearest neighbor in the original 2n, then the classification error of this procedure approaches ε_{nn} as $n\to\infty$ with probability 1. (See [2].)

We now make a final assumption that $p_1 \equiv 1$ on I. Again, results analagous to the following will still hold provided p_1 is strictly bounded away from 0 in I.

Suppose we are given n independent samples from a distribution with density p_2. Let $1\equiv\varphi_1, \varphi_2, \ldots$ be a complete orthonormal system for $L_2(I)$. Finally, let ν_n be the empirical density determined by the n sample points. Now, consider the solution

of the variational problem: minimize

$$\tfrac{1}{2}\,\mathrm{VAR}_1 f + \tfrac{1}{2}\,\mathrm{VAR}_{\nu_n} f = J_N{}^n(f) \qquad (*)$$

such that

$$f = \sum_1^N a_i \varphi_i$$

$$E_1\, f = 0$$

$$E_{\nu_n}\, f = 1$$

where N is determined by a "stopping rule". We then let f_n be the above minimizing f. Before describing the determination of N, we show that the preceding variational problem has solutions with probability 1 for large enough N.

<u>Lemma</u> Assume n is fixed. Then with probability one (*) has solutions for large enough N. In fact $\min_f J_N{}^n(f) \to 0$ as $N \to \infty$ with probability one.

<u>Proof:</u> Let L be any positive integer and $\varepsilon > 0$. Then there is an N_0 such that, for $N > N_0$, there are functions $\Psi_1, \Psi_2, \ldots \Psi_L$ $\varepsilon \langle \varphi_1, \varphi_2, \ldots \varphi_N \rangle$ with the properties:

(i) $\|\Psi_i\|_2^2 \le \frac{1}{L} + \varepsilon$

(ii) there exist disjoint subsets $A_1, \ldots A_2$ with $m(\bigcup_1^L A_i) > 1-\varepsilon$ s.t. $x \in A_i$ implies

$$|\Psi_i(x)-1| < \varepsilon \quad \text{and} \quad |\Psi_j(x)| < \varepsilon \quad (j \ne i)\,.$$

Hence, with probability (wrt p_2) $> (1-K\varepsilon)^n$, each of our samples ℓ will lie in some A_{i_ℓ}. Let us now consider the function

$$\tilde{f} = \frac{\sum_{\ell=1}^{n} \Psi_{i_\ell} - \sum_{\ell=1}^{n} \int \Psi_{i_\ell}\, dx}{\frac{1}{n}\sum_{k=1}^{n} \quad \sum_{\ell=1}^{n} \Psi_{i_\ell}(x_k) - \sum_{\ell=1}^{n} \int \Psi_{i_\ell}\, dx}$$

Clearly $E_1\, \tilde{f}=0$, $E_{\nu_n}\, \tilde{f}=1$. We have further

$$\mathrm{VAR}_1\, \tilde{f} = \|\tilde{f}\|_2^2 \le \left(\frac{n\sqrt{\frac{1}{L} + \varepsilon}}{1-n\varepsilon - n\sqrt{\frac{1}{L} + \varepsilon}} \right)^2$$

$$\mathrm{VAR}_{\nu_n}\, \tilde{f} \le \left(\frac{2n\varepsilon}{1-n\varepsilon - n\sqrt{\frac{1}{L} + \varepsilon}} \right)^2$$

Hence, $J_N^{\,n}(\tilde{f})$ becomes arbitrarily small as $N \to \infty$ with probability arbitrarily close to one.

The solutions of (*) can be easily obtained by the method of Lagrange multipliers. Since $\varphi_1 \equiv 1$, (*) is reduced to solving the following for a_i -

$$\min \left[\tfrac{1}{2} \sum_{2}^{N} a_i^{\,2} + \tfrac{1}{2} \sum_{i,j=2}^{N} a_i a_j \bar{\varphi}_{ij} \right]$$

such that

$$\sum_{2}^{N} a_i \ \bar{\varphi}_i = 1$$

$$\bar{\varphi}_i = \frac{1}{n} \sum_{1}^{n} \varphi_i(x_\ell)$$

$$\bar{\varphi}_{ij} = \frac{1}{n} \sum_{1}^{n} \varphi_i(x_\ell) \ \varphi_j(x_\ell)$$

For the determination of $N=N_n$, we first estimate $J_{\frac{1}{2}}(\bar{f})$ by

$$\bar{J}_n = \tfrac{1}{2} + \frac{\varepsilon^n}{4(\frac{1}{2}-\varepsilon^n)} \tag{5}$$

where ε^n is the expected nearest neighbor error rate of the n samples with the leaving-one-out method:

$$\varepsilon^n = \frac{1}{n} \sum_{\ell=1}^{n} \left[1-(1-V_\ell)^{n-1}\right] \tag{6}$$

where V_ℓ is the volume of the intersection of I with a sphere centered at x_ℓ and of radius equal to the distance between x_ℓ and its nearest neighbor in $\{x_k\}_{k\neq\ell}$. Now $\varepsilon^n \to \varepsilon_{nn}$ with probability one as $n\to\infty$ and hence $\bar{J}_n \to J_{\frac{1}{2}}(\bar{f})$ with probability one as $n\to\infty$. Let N_n be an N which minimizes $|J_N{}^n(f_n) - \bar{J}_n|$. By the lemma such an N exists with probability one provided that $\bar{J}_n > 0$ and this is true with probability one as $n\to\infty$.

The estimate we then use for p_2 is

$$p_n = \frac{f_n - 1 + 2\bar{J}_n}{2\bar{J}_n - f_n} \quad . \tag{7}$$

If we should know the value of K, we may use the truncated estimate

$$\hat{p}_n = (p_n \vee 0) \wedge K \tag{8}$$

We now make the following consistency claim.

<u>Theorem</u> $p_n \to p_2$ in Lebesgue measure with probability one and $\hat{p}_n \xrightarrow{L_2} p_2$ with probability one.

<u>Proof</u> From the form of (3), (7), (8) and the fact that $\bar{J}_n \to J_{\frac{1}{2}}(\bar{f})$, it suffices to show that $f_n \xrightarrow{L_2} \bar{f}$ with probability one.

Note that $\varphi_2, \varphi_3, \ldots$ are linearly independent and dense in $\{f: \int f dx = 0\} \cap L_2(\frac{1}{2}+\frac{p_2}{2})$ where $L_2(\frac{1}{2}+\frac{p_2}{2})$ denotes the set of square integrable functions wrt. to a measure whose density is $\frac{1}{2}+p_2/2$. Now form a complete orthonormal basis $\xi_2, \xi_3, \ldots$ of $\{f: \int f dx = 0\} \cap L_2(\frac{1}{2}+\frac{p_2}{2})$ where each ξ_i is a linear combination of $\varphi_2, \varphi_3, \ldots \varphi_i$. Let $c_i = \int \xi_i \, p_2$. Then $\bar{f} = \sum_2^\infty b_i \xi_i$ where b_i is the solution of $\min \sum_2^\infty b_i^2$ such that $\sum_2^\infty c_i \, b_i = 1$. This is just $b_i = c_i / \sum_2^\infty c_i^2$.

Similarly, we form a complete orthonormal basis $\eta_2{}^n$, $\eta_3{}^n, \ldots$ of $\{f: \int f=0\} \cap L_2(\frac{1}{2}+\frac{\nu_n}{2})$, with each $\eta_i{}^n$ a linear combination of $\varphi_2, \varphi_3, \ldots \varphi_i$. Let $d_i{}^n = \int \eta_i{}^n \ \nu_n$. The solution of (*) is given by $f_n = \sum_2^{N_n} d_i{}^n \ \eta_i{}^n / \sum_2^{N_n} (d_i{}^n)^2$.

Clearly, $d_i{}^n \to c_i$ and $\| \eta_i{}^n - \xi_i \|_2 \to 0$ with probability one as $n \to \infty$ for each i. Since

$$\left| \left(\sum_2^N c_i{}^2 \right)^{-1} - \left(\sum_2^N (d_i{}^n)^2 \right)^{-1} \right| \to 0$$

with probability one as $n \to \infty$, for each N, it follows that

$$\left| \left(\sum_2^{N_n} (d_i{}^n)^2 \right)^{-1} - \left(\sum_2^{\infty} c_i{}^2 \right)^{-1} \right| \to 0$$

with probability one as $n \to \infty$, and hence $\sum_2^{N_n} (d_i{}^n)^2 \to \sum_2^{\infty} c_i{}^2$ with probability one as $n \to \infty$. Finally, for any $\varepsilon > 0$ pick M such that $\sum_M^{\infty} c_i{}^2 < \varepsilon$. Then

$$\overline{\lim} \ \| f_n - \bar{f} \|_2 \le \overline{\lim} \ \| \sum_M^{N_n} d_i{}^n \ \eta_i{}^n \|_2 + \| \sum_M^{\infty} c_i \ \xi_i \|_2$$

with probability one. But

$$\| g \|_2 \le \sqrt{2} \ \| g \|_{\frac{1}{2}+\frac{P_2}{2}} \quad \text{and} \quad \| g \|_2 \le \sqrt{2} \ \| g \|_{\frac{1}{2}+\frac{\nu_n}{2}} \ .$$

Hence $\overline{\lim}\ \|f_n-\bar{f}\|_2 \le 2\sqrt{2\varepsilon}$ with probability one.
Since ε was arbitrary, the proof is complete.

IV. A REMARK ON HYPOTHESIS TESTING

If we are given two sets of data $\{x_i\}_{i=1}^n$ and $\{y_j\}_{j=1}^n$ and are then given the task of constructing an optimal discriminant between the two classes, we might solve

$$\min \tfrac{1}{2}\left[\mathrm{VAR}_{\mu_n} + \tfrac{1}{2}\,\mathrm{VAR}_{\nu_n}\right] \qquad (**)$$

such that

$$E_{\mu_n} f=0$$

$$E_{\nu_n} f=1$$

$$f = \sum_1^{N_n} \varphi_i$$

μ_n, ν_n empirical densities for $\{x_i\}$, $\{y_j\}$

where N_n is determined by the analagous "stopping rule".

Unfortunately the consistency proof does not apply in this case since we are unable to find a bound for $\|g\|_{\frac{p_1+p_2}{2}}$ in terms of $\|g\|_{\frac{\nu_n+\mu_n}{2}}$. Hence, the reference density $p_1 \equiv 1$ "forced" the consistency. We therefore recommend the estimates: $\hat{p}_1$ by the method of III using $\{x_i\}$ and then similarly $\hat{p}_2$ using $\{y_j\}$. The discriminant $\hat{p}_2/\hat{p}_1$ will then be optimal with probability one as $n\to\infty$.

REFERENCES

[1]. L. Jones, "K^{th} Order Solutions to the Problem of Finding Optimal Discriminant Functions", Technical Note 1980-4, Lincoln Laboratory, M.I.T. (10 January 1980), submitted to S.I.A.M. J. Appl. Math. DTIC AD-A0822396/3.

[2]. T. Cover and P. Hart, "Nearest Neighbor Pattern Classification", IEEE Trans. Inf. Theory IT-13, pgs. 21-27 (1967).

(14) TN-1980-54

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(19) **REPORT DOCUMENTATION PAGE**

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
(17) ESD-TR-80-230

2. GOVT ACCESSION NO.
AD A094728

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
(6) On Nonparametric Probability Density Estimation Using Orthogonal Series.

5. TYPE OF REPORT & PERIOD COVERED
(9) Technical Note

6. PERFORMING ORG. REPORT NUMBER
Technical Note 1980-54

7. AUTHOR(s)
(10) Lee K. Jones

8. CONTRACT OR GRANT NUMBER(s)
(15) F19628-80-C-0002

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Lincoln Laboratory, M.I.T.
P.O. Box 73
Lexington, MA 02173
(16) 627A

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Program Element No. 63311F
Project No. 627A

11. CONTROLLING OFFICE NAME AND ADDRESS
Air Force Systems Command, USAF
Andrews AFB
Washington, DC 20331

12. REPORT DATE
(11) 26 Nov~~ember 1~~980

13. NUMBER OF PAGES
16

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
Electronic Systems Division
Hanscom AFB
Bedford, MA 01731
(12) 15

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

density estimation — Fourier coefficients — orthogonal expansions

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A method of density estimation is proposed, which is a rational modification of orthogonal expansions, combined with a stopping rule determined by a nearest neighbor statistic. This method yields consistent estimates and applies (in principle) to density estimation in any number of dimensions.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

201650 JU

LMED

8